Reconstruction of a genome-scale metabolic model for the filamentous fungus *Ashbya gossypii*

Daniel Gomes, Oscar Dias, Eugénio Ferreira, Lucília Domingues, Isabel Rocha

IBB – Institute for Biotechnology and Bioengineering, Centre of Biological Engineering

Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal

E-mail {danielg_gomes, odias, ecferreira, luciliad, irocha}@deb.uminho.pt

Abstract

Systems biology has recently arisen as a promising and powerful tool for process development and optimization. Metabolic models are one of its different methodologies with high interest and applicability since it allows the simulation of cells behavior under different environments and/or specific genetic variations. The fast-growing number of sequenced genomes may have contributed to this phenomenon, as the sequenced genome is the starting point from where it is possible to associate by homology a specific function to the genes of a microorganism. Afterwards, from this entire set of enzymatic functions we can possibly identify the main metabolic pathways of a specific microorganism allowing this way the construction of its metabolic network.

One of the genomes already sequenced is from the fungus *Ashbya gossypii*, an industrially relevant microorganism intensively used for riboflavin production. Despite the high similarity with *Saccharomyces cerevisiae* genome *A. gossypii* presents a lower level of complexity containing only 4726 protein-coding genes distributed over seven chromosomes. The aim of this work is to construct a metabolic model for *A. gossypii* based on its genome and from which we can retrieve valuable information concerning specific metabolic pathways and the optimum conditions for the production of interesting compounds such as riboflavin.

The initial stage of this process consisted in the collection of all metabolic-relevant genes through a manual re-annotation of *A. gossypii* genome. Despite being a manual procedure, this step was performed using the user-friendly software – *merlin* – that provided an automatic annotation for each gene, speeding up the entire process. The function automatically assigned by this application was manually analyzed, being accepted or replaced by another one. Each metabolic gene was assigned to an Enzyme Commission(EC) number that corresponds to a specific enzyme. For such procedure several databases were used such as UniProt, SGD, AGD, ExPASy and BRENDA.

At the end of this phase, a total of 1429 genes were assigned among the different enzymatic families. Such distribution was considerably heterogeneous: 35,4 % hydrolases; 35,8 % transferases; 28,8 % other enzymatic families. Of the 1429 genes 59 were assigned to different EC numbers and among these 36 % have EC numbers from different enzymatic families.

The next stage of the reconstruction process, being performed at present, involves the elaboration of the set of metabolic reactions and their curation regarding stoichiometry, balance of charges and localization inside the cell. For this purpose, information from genome re-annotation is crossed with curated models from closely related microorganisms such as the iMM904 (Mo et al., 2009) and the iIN800 (Nookaew et al., 2008) from *Saccharomyces cerevisiae*. To complete this process regarding information that was not found in the curated models, reactions databases like BRENDA or KEGG will be used to retrieve such data.

At the end of this phase, once we have the complete set of curated metabolic reactions, we will be able to construct a system of m equations (metabolites) and n variables (reactions) that is the base for the optimization studies. The next step will be the determination of biomass composition that encounters another key element for the optimization studies.

Acknowledgements: The authors thank the financial support of Fundaçãopara a Ciência e aTecnologia (FCT), Portugal: project AshByofactoryPTDC/EBB-EBI/101985/2008.